LA-UR-21-26296
Approved for public release; distribution is unlimited.

Title: LANL Platforms Update

Author(s): Lujan, James Westley

Intended for: International Conference

Issued: 2021-07-09 (rev.1)

HIGH PERFORMANCE COMPUTING

Los Alamos National Laboratory

LANL Platforms Update

Jim Luján

July 12, 2021

LA-UR-26296

LANL Platforms

Current Capability / Commodity Systems

- Fire / Ice / Cyclone (CTS-1 – 1.33 PF/s, 1104 nodes Intel Broadwell)
- Snow (CTS-1 – 445 TF/s, 368 nodes Intel Broadwell)
- Grizzly (1490 nodes 1.806 PF/s) / Badger (660 nodes 798 TF/s) Intel Broadwell
- Chicoma (1024 nodes) AMD Rome

Advanced Technology Systems

- Trinity (ATS-1 – 40 PF/s – 9408 Intel Haswell / 9800 KNL)
- Trinitite (364 TF/s – 100 Intel Haswell / 100 KNL)

Application Testbeds

- Capulin / Thunder –Cray, Arm TX2 system (167 nodes)
- Hinata / Akebono – Fujitsu A64FX – 32 processors

Future Systems

- NVidia SDK testbeds (Arm / A100 - TBD) FY21
- Chicoma+ (NVidia A100) FY21 / NGP-1 (NVidia Arm / A100+) FY23
- CTS-2 – FY22 / Crossroads (ATS-3) – FY22 / ATS-5 – FY26/27

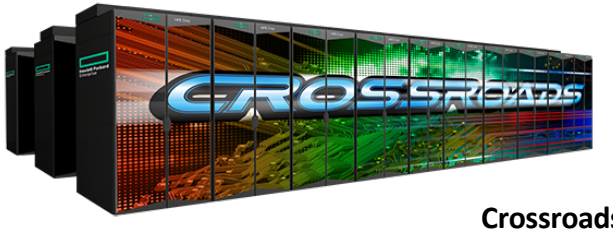
Chicoma (IC)

- HPE Cray EX (12 compute cabinets with 3 CDUs)
  - Similar to Crossroads, Perlmutter, Frontier, etc.
    - Multi-architectural / multi-generational racks allow for future expansion
- 2 racks currently populated with AMD Rome blades
  - 2 racks x 64 blades x 4 dual-socket nodes @ 128 cores per socket
    ~72,000 cores (~54,000 in Grizzly)
  - 512 GB DDR4 per node, 8 channels per socket (256 TB)
  - Cray Slingshot interconnect, Cray CSM, Cray PE
- Future additions:
  - 118 GPU nodes (at least): single socket AMD Milan + 4 NVIDIA A100
  - Additional purchases in FY22 and beyond

Crossroads (ASC)

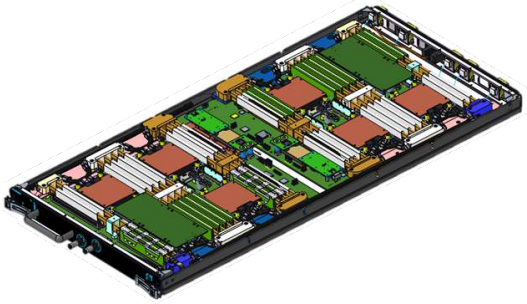
- Crossroads was awarded March 2020
  - HPE / Cray were integrating
    - HPE Apollo v Cray Shasta Packaging
    - HPCM software v Cray Shasta Software
    - IB v Slingshot
    - Marvell TX4 based system
- Designed contract in anticipation of HPE/Cray integration – TDPs
  - Processor, I/O, Software, Interconnect, etc.
- Then…
  - Marvell quietly dropped HPC line of Thunder processors
  - Transitioned to alternate technology proposal – Intel SPR w/advanced memory

Crossroads System Overview



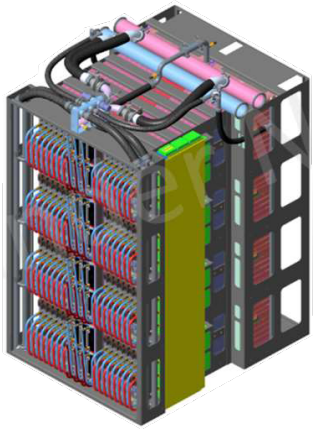
Crossroads System

24	Cray EX Cabinets (Olympus)
8	CDUs
1,536	Compute Trays
6,144	Dual-socket nodes
12,288	Intel Xeon Sapphire Rapids
768	Rosetta Switches



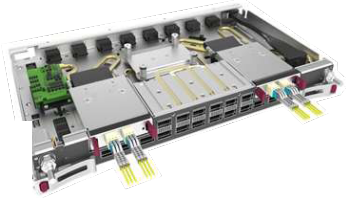
Compute Tray

2	Compute boards
4	Dual-socket nodes
8	Intel Xeon Sapphire Rapids
8	Cassini NICs



Cray EX Cabinet (Olympus)

8	Chassis
64	Compute Trays
256	Dual-socket nodes
512	Intel Xeon Sapphire Rapids
32	Fabric Trays
32	Rosetta switches



Fabric Tray

1	Compute boards
16	Dual-socket nodes
12	HPE Stringer optics ports
36	Pluggable ports

Crossroads – V? (Intel delays SPR/HBM) – And now…

- Install a substantial portion of Crossroads earlier but with DDR, then upgrade it when HBM parts are available.
- Installation of first half (Phase 1 w/DDR) in early summer 2022
  - Accept Phase 1 (functionality, performance, stability)
  - Transition into the Secure for production cycles
- Installation of second half (Phase 2 w/HBM) late 2022
  - Accept Phase 2 (functionality, performance, stability)
  - Transition into the Secure for production cycles
- Upgrade Phase 1 early in 2023.
  - Remove all DDR SPR sockets and replace with SPR/HBM
  - Crossroads 100% SPR/HBM (6,144 nodes)
- Complete application performance acceptance (full system)

Upgrade 3K SPR nodes w/ 256GB DDR to SPR+HBM

	Phase 1 Install SPR/DDR Compute Nodes	Phase 2 Install SPR/HBM2e Compute Nodes	Phase 3 Upgrade Phase 1 to SPR/HBM2e	Final Config
Installation	**May 2022**	**November 2022**	**Spring 2023**	**Spring 2023**
Node Config	2 x 56C 2.3Ghz SPR, 16 x 16GB DDR5 4800	2 x 56C 2.3Ghz SPR+HBM each w/ 64GB of HBM2e	2 x 56C 2.3Ghz SPR+HBM each w/ 64GB of HBM2e	
Node Count	**3072**	**3072**	**3072**	**6144**
DDR GB/Node	**256**	**0**	**0**	
HBM GB/Node		**128**	**128**	**128**
Gflops / Node	**8,243**	**8,243**	**8,243**	
Total DDR PiB	**0.876**			
Total HBM PiB		**0.393**	**0.393**	**0.786**
Total Peak PF (DP)	**25.3**	**25.3**	**25.3**	**50.6**

Crossroads - Ongoing Risk Mitigation

- Intel has to provide regular (monthly) updates to demonstrate significant progress, milestones completed towards SPR/HBM delivery (rebuild trust)
- Intel to recommit to SPR/HBM product line (new CEO)
- Crossroads contract is renegotiated with updated milestone schedule to include a final technology commitment (late summer of 2021)

Opportunity to repurpose / retain 6,144 SPR Xeon SPR processors and 49,152 16GB DDR5 DIMMs (786TB)

NGP-1 – Forward looking

- $80M, delivery in early FY23
- CoDesign^2 innovation; focus on advancing methods for complex Multiphysics, Machine Learning, Analytics…
- Arm is a powerful chance to shape the future
  - Nimble and affordable tailoring for special needs
  - Other countries have demonstrated great value [e.g Fugaku]
- Modest sized system
- HPE Integrated / Shasta / Slingshot
- Mixed processing
  - CPU only (Grace, Arm-based) ~25%
  - CPU (Grace) / GPU (next-gen) ~75%
- Other details (NDA)

NGP-1 Resource Goals

- Significant investment for computing at LANL
- Unclassified initially (N years) then in secure after that
- Serve the institution including weapons and open science uses
- Not IC or ASC owned – for the institution
- Stated uses drive use governance:
  - Institutional directional computing >=1/3
  - Weapons/ASC >=1/3
  - Significant time is allotted to NGP related activities (not application science but computing science (try new system level solution space to assist with tailoring evaluation, etc. pointed at ATS-5 bids/NRE) <=1/3